ImplicitAtlas: Learning Deformable Shape Templates in Medical Imaging

Jiancheng Yang[1,2] Udaranga Wickramasinghe[2] Bingbing Ni[1*] Pascal Fua[2]
[1]Shanghai Jiao Tong University, Shanghai, China [2]EPFL, Lausanne, Switzerland

Abstract

Deep implicit shape models have become popular in the computer vision community at large but less so for biomedical applications. This is in part because large training databases do not exist and in part because biomedical annotations are often noisy. In this paper, we show that by introducing templates within the deep learning pipeline we can overcome these problems. The proposed framework, named ImplicitAtlas, represents a shape as a deformation field from a learned template field, where multiple templates could be integrated to improve the shape representation capacity at negligible computational cost. Extensive experiments on three medical shape datasets prove the superiority over current implicit representation methods.

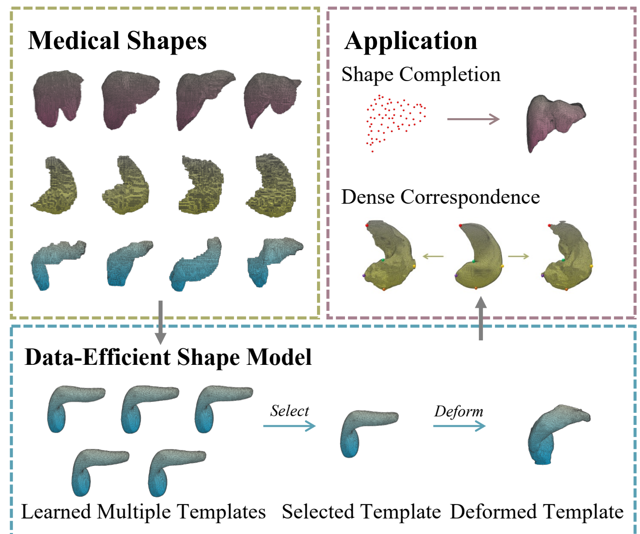

Figure 1. ***ImplicitAtlas*.** We propose a data-efficient shape model for medical imaging. It represents shapes by deforming one of several learned templates.

1. Introduction

Shape modeling is central to medical image analysis, and many different approaches to surface representation have been used for this purpose [16, 36]. In recent years, deep implicit surfaces [8, 38, 44] have emerged as a powerful alternative to more established methods. This is particularly true in the field of computer vision at large but less so in the subfield of biomedical imaging.

This is attributable to the specific challenges posed by biomedical datasets [27, 56]: For manufactured objects, there are large datasets [6, 31] that can be used for training purposes of shape models. These do not exist for many objects of interests in medical imaging, such as the shape of organs and lesions. Even when medical images and the corresponding 3D models are available, the models are of much lower quality due to the complexity and expense of precise annotation. Furthermore, within a single dataset, spatial resolution is rarely constant and often anisotropic. Human errors can result in labeling noise, and organ boundaries are often cropped due to the limitations of imaging process. Fig. 2 illustrates some of these difficulties.

To address these issues, we propose *ImplicitAtlas*, a data-efficient implicit shape model for medical imaging. During training, exemplars represented on discrete grids are taken as input, the most common representation in the biomedical imaging, and the model outputs a continuous occupancy grid. Here, an implicit function learns **multiple** templates, which can undergo non-rigid deformations learned by another implicit function. As in multi-atlas segmentation [21, 58], the templates make our approach better at dealing with limited training data and less sensitive to label noise. Thanks to a straight-through estimator (STE) [3], multiple templates can be learned in an end-to-end fashion at negligible computational cost. Finally, to improve the data efficiency further, we introduce a convolutional implicit function [9, 45] to extract multi-scale features.

To demonstrate the effectiveness of *ImplicitAtlas*, we perform extensive experiments on three medical shape datasets of liver, hippocampus and pancreas. Our method outperforms current implicit representation-based methods [12, 38, 45, 72] by considerable margins, especially when trained in a few-shot learning setting. We also demonstrate several potential applications, such as shape completion from user-supplied point annotations and keypoint la-

*Corresponding author.

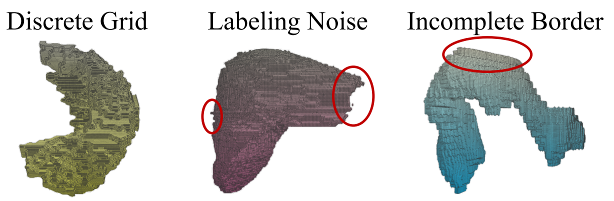

Figure 2. **Common Artefacts in Biomedical Shapes.** As medical shapes are annotated in image stacks, they are generally represented in *discrete grids*. The annotations usually include *label noise* and *incomplete borders*.

beling by the learned dense correspondences. As will be shown, in spite of the challenges that biomedical datasets pose, the proposed implicit methods is made very effective by allowing them to learn multiple templates from training data, and to choose one to reconstruct a particular organ.

2. Related Work

CNN-based volumetric methods such as U-Net and its variants [10, 22, 40, 53] now dominate biomedical image segmentation. This is evident from the CHAOS challenge [29] and Medical Segmentation Decathlon [24] results. The winners of both competitions used ensembles of methods relying on volumetric CNNs to handle the traditional volumetric image segmentation problem. However, biomedical image segmentation is far from solved in practice. For instance, it is still difficult to preserve geometric and topological structures. Meanwhile, volumes can be recovered accurately: obtaining high-quality surfaces remains hard [34].

These challenges highlight the importance of shape modeling in medical image analysis [16, 36], including medical image segmentation [33, 50, 55, 59], computer-aided diagnosis [4, 59, 71], and computational anatomy [5].

Explicit Representations. As shapes in medical imaging are generally annotated on discrete grids from image stacks—acquired by computed tomography (CT) and magnetic resonance imaging (MRI), among others—-most prior art [33, 50, 55, 59, 64] relies on voxel representations. As a 3D extension of 2D pixel grid, voxel neural networks could be developed by extending the corresponding 2D versions (*e.g.*, 3D UNet [11]) with 3D convolutions, or using sophisticated operators [67–69]. Unfortunately, they have high-memory requirements for a relatively low spatial resolution. Hence, the use of geometric data structures, such as point clouds and meshes, have also been explored. Point clouds are lightweight and flexible in sensing and processing [7, 23, 48, 70]. They are suitable to extract semantic information [18, 19, 66] but do not capture topology. Recovering the surfaces from point clouds is a non-trivial problem, which requires refined techniques [15, 41]. Triangulated meshes allow memory-efficient processing for high-fidelity surface reconstruction [14, 28, 61, 62] but changing their topology is non-trivial. There are algorithms designed for this purpose [37], but they require *ad hoc* heuristics that do not generalize well.

Implicit Representations. Recently, implicit representations [8, 38, 44] has become increasingly popular in deep learning-based 3D computer vision. They represent a 3D shape as an isosurface in a continuous 3D field, which is parameterized by a deep network. Due to their flexibility, memory efficiency, and ability to represent any topology at any resolution, implicit representations have been widely investigated not only for shape, but also appearance [39] and scene [42] modeling.

However, they have not yet made big inroads in medical image analysis. On of the few studies in that area can be found in [49]. But it focuses on refining medical image segmentation produced using existing implicit representation methods. We instead focus on a high-quality implicit representation method to address the difficulty of developing implicit fields in biomedical imaging area.

This paucity of implicit methods can be attributed to the specific challenges in biomedical image processing: Large databases are rarely available and annotations are often imprecise. Using atlases and templates in shape modeling has long been known as a good approach to tackle these issues, and we discuss them below.

Atlases and Templates. Probabilistic atlases are widely used for atlas-based image segmentation [21, 58] because they are an excellent mean to deal with the noise in biomedical imagery. With the advent of deep learning, atlases have been integrated into convolutional neural networks [2, 13, 20, 54]. All these approaches rely on pre-computed atlases. They are created by fusing multiple manually annotated images; the atlases must also be pre-registered to the target images to align them with the structures of interest. In [60], an attempt is made to use atlases that could be automatically aligned and deformed to match the target structures.

On the other hand, templates are used in conjunction with implicit surfaces in [12, 72]. This involves using an implicit method to predict deformations around a template, where the deformation and template are both parameterized implicitly. However, these methods were developed on large training datasets in mind, and data efficiency is not the primary focus. Both rely on MLP decoders that do not introduce spatial reductive bias as convolutional ones do [45]. Besides, only a single implicit template can be automatically learned by these methods, and a central argument of this paper is that it is beneficial to more than one.

3. Methodology

In this section, we first briefly review deep implicit shape representations. We then introduce our model, the corresponding architecture, and our training approach.

3.1. Background: Deep Implicit Surfaces

Implicit shape representations [8, 38, 44] model shapes by mapping 3D coordinates to a shape indicator, typically occupancy or signed / unsigned distance. In this work, we use the latter. For a shape S, this mapping is expressed as

$$\mathcal{F}(\mathbf{h}, \mathbf{p}) = o : \mathbb{R}^c \times \mathbb{R}^3 \to \mathbb{R} \,, \tag{1}$$

where $\mathbf{h}$ is a c-dimension *latent vector* that encodes S, $\mathbf{p}(x, y, z) \in \mathbb{R}^3$ is a query point, and $\mathcal{F}$ is implemented by a deep network that outputs the probability of occupancy $o \in [0, 1]$. o should be close to 1 for $\mathbf{p}$ inside S and 0 otherwise. Given a training set, $\mathcal{F}$ and the vector $\mathbf{h}$ corresponding to each shape can be learned in many ways. Here, we use an auto-decoding approach as in [38], where the $\mathbf{h}$ is treated as a learnable parameter and jointly optimized with the parameters of $\mathcal{F}$.

Apart from the methods that directly output shape indicators, there are also studies that regard shapes as deformation from templates [12,72]. $\mathcal{F}$ is a rewritten as a composite function of $\mathcal{T}$ and $\mathcal{D}$, that is,

$$\mathcal{F}(\mathbf{h}, \mathbf{p}) = \mathcal{T}(\mathcal{D}(\mathbf{h}, \mathbf{p})) \,, \tag{2}$$

where $\mathcal{D} : \mathbb{R}^c \times \mathbb{R}^3 \to \mathbb{R}^3$ is a function that maps a query point $\mathbf{p}$ to new coordinates $\mathbf{p}'$ and $\mathcal{T} : \mathbb{R}^3 \to \mathbb{R}$ is a learned implicit function that plays the same role as $\mathcal{F}$ in Eq. 1 but learns a single shape. $\mathcal{D}$ can be implemented in several ways, including an additive deformation [12] or a point-wise affine transformation [72]. In this formulation, $\mathcal{T}$ plays the role of a template because it encodes a shape prior learned from the training shapes. Notably, as all the deformed coordinates are aligned with the template, dense correspondences between shapes can easily be established.

3.2. Model Pipeline

Although earlier than the deep learning era, multi-atlas techniques [21, 58] are good at dealing with limited data and label noise. We translate them into our framework as depicted by Fig. 3 (a). Given the formulation of Eq. 2, we take $\mathcal{T}$ to be an *Implicit Template Network* ($\mathcal{T}$) and $\mathcal{D}$ to be an *Implicit Deformation Network*, Eq. 2 is rewritten as

$$\mathcal{F}(\mathbf{h}, \mathbf{p}) = \mathcal{T}(\mathbf{t}(\mathbf{h}), \mathbf{p} + \mathcal{D}(\mathbf{d}(\mathbf{h}), \mathbf{p})) \,, \tag{3}$$

where $\mathbf{t}$ and $\mathbf{d}$ are separate vectors that are functions of $\mathbf{h}$, and the output of $\mathcal{D}$ is the deformation (instead of the deformed coordinates in Eq. 2). Note that $\mathcal{T}$ now has an additional argument, which we are going to use select one template among several possible ones. This design enables the learning of multiple templates and improves the representation capacity of our model over earlier work [12, 72].

Template Selection. We introduce a matrix of learnable parameters $\mathbf{T} \in \mathbb{R}^{m \times c}$, where m denotes the number of templates and c is the dimension of the latent vectors. Template selection is achieved by picking a row vector $\mathbf{t}$ from $\mathbf{T}$ and feeding to the implicit function ($\mathcal{T}$). $\mathbf{t}$ is taken to be

$$\begin{aligned} \mathbf{t} &= \textit{STE-Softmax}(\mathbf{s}) \cdot \mathbf{T}, \\ \mathbf{s} &= \textit{MLP}(\mathbf{h}) \in \mathbb{R}^m, \end{aligned} \tag{4}$$

where STE-Softmax is a softmax with straight-through estimator [3]: the softmax is "hardened" as a one-hot in the forward pass, but it directly takes the gradient of the one-hot in the backward pass. This can be regarded as *reparameterization* [30] for categorical distributions. The STE-Softmax could be replaced by Gumbel-Softmax [26, 35], which, in theory, should provide a smoother gradient for categorical reparameterization. However, in our experiments, it does not make a difference.

Estimating the Deformation. The variable $\mathbf{d}$ of Eq. 3 controls the deformation at the query point $\mathbf{p}$, which we take to be a point-wise additive deformation: $\mathbf{p} \to \mathbf{p} + \mathcal{D}(\mathbf{d}(\mathbf{h}), \mathbf{p})$. As medical shapes are represented in discrete grids, $\mathbf{p}$ is implemented as a meshgrid $\mathbf{P} \in \mathbb{R}^{D \times H \times W \times 3}$, where $D \times H \times W$ denotes the spatial size.

3.3. Network Architecture

Multi-Layer Perceptrons (MLPs) have been a popular choice to parameterize the deep implicit functions [8, 12, 38, 44, 72]. Unfortunately, these MLPs tend to be data hungry. We therefore used a convolutional decoder instead, as in ConvONet [45].

As illustrated by Fig. 3 (b), a latent vector—either $\mathbf{t}$ or $\mathbf{d}$ in our approach—is first transformed into multi-scale feature maps by convolutional and upsampling layers, from which the query point $\mathbf{p}$ obtains its features as a function of its coordinates via trilinear interpolation [25]. Instead of only using the final feature map as in [45], we extract multi-scale features from the feature maps at several resolutions as in [9]. Finally, the coordinates along with the resulting features encoding local and global semantic information are concatenated and fed into a small MLP to produce an output. The multi-scale features make the model less data-hungry than pure MLPs.

$\mathcal{T}$ and $\mathcal{D}$ are implemented using the same convolutional decoder, except for the final layer. There is one output channel for $\mathcal{T}$ and three for $\mathcal{D}$. Each convolutional block is a stack of convolution layer, group normalization [63] and leaky ReLU activation [65]. The first upsampling layer is implemented as pixel shuffle [52]. More implementation details can be found in supplementary materials.

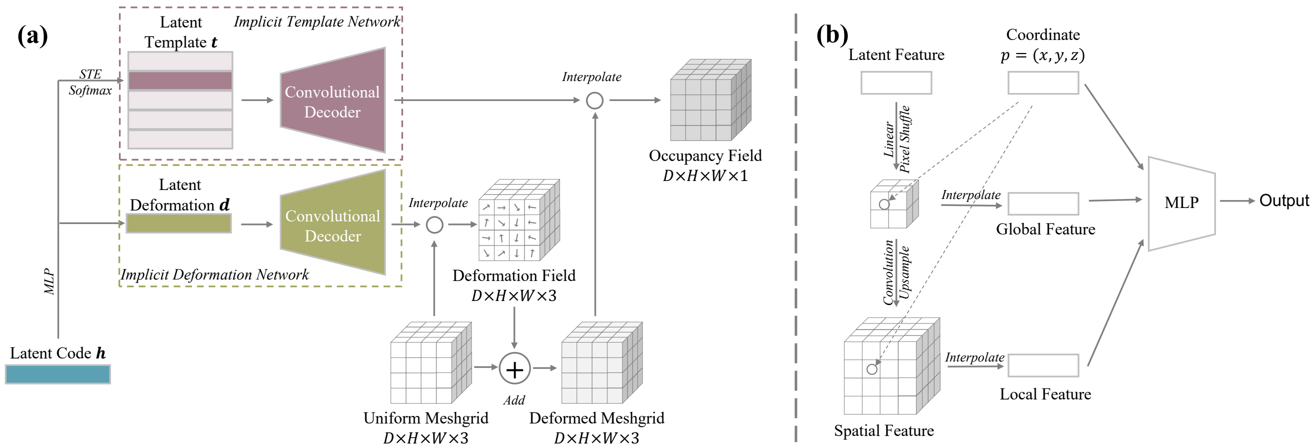

Figure 3. **Overview of *ImplicitAtlas*. (a) Model Pipeline.** The model consists of an *Implicit Template Network* ($\mathcal{T}$) and an *Implicit Deformation Network* ($\mathcal{D}$). Given a latent code $\mathbf{h}$, it selects a latent template $\mathbf{t}$ via STE-Softmax to generate a template using $\mathcal{T}$. It also produces a latent deformation $\mathbf{d}$ to generate a deformation field from the template using $\mathcal{D}$. They are composed to produce an occupancy field, which is the final output. **(b) Network Architecture of the Decoder.** Given a latent feature, it builds multi-scale feature maps in a convolutional manner. For a query coordinate $\mathbf{p} = (x, y, z)$, it aggregates both local and global feature by interpolating on the multi-scale feature maps. Finally, the coordinate and the interpolated features are fed into an MLP for final output.

3.4. Model Training

As we take occupancy formulation, the primary task loss to minimize is the binary cross entropy,

$$\mathcal{L}_{Task} = -\frac{1}{N}\sum \hat{o}\cdot log(o) + (1-\hat{o})\cdot log(1-o)\ , \quad (5)$$

where N is the number of sampled points, o and $\hat{o}$ are the predicted and true occupancy, respectively. Given that we use an auto-decoding approach [44] to learning the latent vectors and using that these latent codes from a Gaussian prior distribution, we add the regularization penalty

$$\mathcal{L}_2 = ||\mathbf{h}||_2\ , \quad (6)$$

where $||\cdot||_2$ denotes the l_2-norm. To further regularize the model when training with limited data, we define 2 additional regularization terms: Laplacian Smoothness ($\mathcal{L}_{LS}$) and a Deformation Penalty ($\mathcal{L}_{DP}$), written as:

$$\mathcal{L}_{LS} = \frac{1}{N}\sum \sum_{w\in\{x,y,z\}} ||\mathcal{F}_{w+1} + \mathcal{F}_{w-1} - 2\mathcal{F}_w||_2\ ,$$

$$\mathcal{L}_{DP} = \max ||\mathcal{D}||_2 + \frac{1}{N}\sum ||\mathcal{D}||_2\ , \quad (7)$$

where $\mathcal{F}_x$ is short for $\mathcal{F}(\mathbf{h}, (x, y, z))$—and similarly for y and z—and $\mathcal{D}$ is short for $\mathcal{D}(\mathbf{h}, \mathbf{p})$. Minimizing $\mathcal{L}_{LS}$ favors a spatially smooth output while minimizing $\mathcal{L}_{DP}$ restricts the deformations so that $\mathcal{T}$ has to learn more details.

During training, we uniformly sample a meshgrid $\tilde{P} \in \mathbb{R}^{\tilde{D}\times\tilde{H}\times\tilde{W}\times 3}$ of lower resolution than the volume of interest. For the shapes of 128^3 in this study, we sample 32^3 meshgrid during training. Each point in the meshgrid is added with a random noise, and its occupancy ground truth is sampled in the full-resolution volume of interest. This sub-griding technique significant reduces the training cost in terms of both memory and time. Furthermore, as we sample query points on a uniform grid, computing $\mathcal{L}_{LS}$ can be done efficiently using a 3D convolution with a custom kernel. $\mathcal{L}_{DP}$ is calculated over the meshgrid.

Organ	# Known	# Unkown	Annotation Modality
Liver	111	20	Portal venous phase CT
Hippocampus	221	39	Mono-modal MRI
Pancreas	238	43	Portal venous phase CT

Table 1. **Datasets for the Liver, Hippocampus and Pancreas.** Known (training) and unknown (test) shapes are based on the organ segmentations from the Medical Segmentation Decathlon [1].

4. Experiments

We now evaluate our *ImplicitAtlas* approach both quantitatively and qualitatively.

4.1. Datasets

For our experiments, we use the Medical Segmentation Decathlon (MSD) [1], which is arguably the largest and most comprehensive medical image segmentation data set available to date. It is a collection of 10 medical image datasets, featuring several organs imaged using many different modalities. To test our approach under many different conditions, we experimented with three organs that pose different challenges: the liver, a big organ with complex de-

Method	Liver (K) DSC	Liver (K) NSD	Hippocampus (K) DSC	Hippocampus (K) NSD	Pancreas (K) DSC	Pancreas (K) NSD	Liver (U) DSC	Liver (U) NSD	Hippocampus (U) DSC	Hippocampus (U) NSD	Pancreas (U) DSC	Pancreas (U) NSD
MLP Decoder [38, 44]	96.32	95.20	93.21	91.50	94.83	95.55	93.12	71.23	90.20	63.05	89.46	65.15
+ Template [12, 72]	97.77	97.65	93.99	92.00	95.89	96.54	94.26	81.28	91.89	67.92	90.32	71.65
Conv Decoder [9, 45]	98.47	98.19	94.88	92.64	95.54	96.18	96.61	86.12	93.27	75.87	92.22	80.14
ImplicitAtlas	**98.58**	**98.69**	**96.42**	**94.72**	**96.85**	**97.30**	96.59	85.95	93.54	76.99	**93.38**	**81.11**
ImplicitAtlas + reg.	98.50	98.33	96.09	92.85	96.76	97.03	**96.72**	**86.90**	**93.99**	**77.47**	93.31	80.92

Table 2. **Reconstruction Accuracy on Known (K) and Unknown (U) Shapes.** The accuracy is expressed in terms of the DSC (%, ↑) and NSD (%, ↑) metrics. reg.: regularization with Laplacian Smoothness and Deformation Penalty.

tails; the hippocampus, a small organ; the pancreas, a soft organ that can be seen in diverse poses.

As annotations are available only for the official training split, we split the data into known (training) and unknown (test) shapes for each organ. All the organ annotations are cropped using the bounding-boxes of organ segmentation, and resized into a fixed size of 128^3 via spline interpolation. Tab. 1 summarizes the basic information about the 3 resulting datasets.

4.2. Baseline Approaches

Most current shape representation methods are trained on CAD models or scenes, rather than on biomedical shapes. For a fair comparison, we reimplemented them and trained them on our datasets. The size of the networks were chosen to be similar to that we use. We use the same training and inference procedures in all cases. The following baseline methods are implemented:

MLP Decoder. Due to the simplicity and flexibility of MLPs, many algorithms [8, 38, 44] use them as decoders. Our MLP baseline uses the architecture of DeepSDF [44].

MLP Decoder + Template. Recent algorithms that rely on templates [12, 72] use different MLP-based network architectures and deformation formulation. To make the comparison meaningful, we implemented a baseline based on Eq. 3 in which $\mathcal{T}$ and $\mathcal{D}$ are implemented with the same MLP architecture, which duplicates what is done in [12].

Conv Decoder. The problem is the same with earlier methods that rely on convolutional implicit fields [9, 45]. To compare, we implemented a baseline based on Eq. 1 that uses the same architecture of our decoder.

4.3. Representing Known and Unknown Shapes

We first present our experimental procedure, and then analyze the reconstruction results and ablation studies.

Experimental Setting. As in earlier work [12, 44, 72], we first evaluate the representation capacity of our model for encoding known and unknown shapes. We train three separate models for the liver, hippocampus, and pancreas. We then evaluate the quality of reconstructions on the training and testing set separately. The first is regarded as the set of the *known* (K) shapes and the second as the *unknown* (U). For the former, the reconstruction is performed during the training of auto-decoding. For the latter, we optimize randomly initialized latent code $\mathbf{h}$ to reconstruct the shapes with fixed model weights.

Unless otherwise specified, we use m=5 templates in our method. A group of fixed loss weights is used for all cases: 10^{-3} for $\mathcal{L}_2$, 10^{-3} for $\mathcal{L}_{LS}$ and 10^{-2} for $\mathcal{L}_{DP}$ We give additional details on the training and reconstruction procedures in the supplementary material.

Metrics. To evaluate the shape similarity between the ground truth and predicted reconstruction, we use Dice Similarity Coefficient (DSC) and Normalized Surface Dice (NSD) [43], which are the standard metrics used in the medical imaging literature. They are defined as

$$DSC(\mathbf{O}, \hat{\mathbf{O}}) = \frac{2|\mathbf{O} \cap \hat{\mathbf{O}}|}{|\mathbf{O}| + |\hat{\mathbf{O}}|}, \tag{8}$$

$$NSD(\mathbf{O}, \hat{\mathbf{O}}) = \frac{|\partial\mathbf{O} \cap B^{(\tau)}_{\partial\hat{\mathbf{O}}}| + |\partial\hat{\mathbf{O}} \cap B^{(\tau)}_{\partial\mathbf{O}}|}{|\partial\mathbf{O}| + |\partial\hat{\mathbf{O}}|}, \tag{9}$$

where $\mathbf{O} \in \mathbb{R}^{D\times H\times W}$ and $\hat{\mathbf{O}} \in \mathbb{R}^{D\times H\times W}$ are the predicted reconstruction and ground truth of the volume, respectively, and $B^{(\tau)}_{\partial V} = \{x \in \mathbb{R}^3 \,|\, \exists\tilde{x} \in \partial V, ||x - \tilde{x}||_2 \leq \tau\}$ denotes the border region of surface ∂V at tolerance τ. Here we take $\tau = 1$. Conceptually, both DSC and NSD are IoU-like metrics in [0,1] (higher is better): DSC measures the volume overlap, while NSD measures the surface overlap.

Results. We compare two versions of *ImplicitAtlas*, with and without the regularization terms of Eq. 7, to the baselines. The results are reported in Tab. 2. The pure MLP decoder underperforms other methods, especially in surface-based measure (NSD) but it can be improved by using the template-based method. The convolutional decoder significantly boosts the shape representation capacity, which indicates the importance of the spatial reductive bias for medical shapes. Our approach that relies on multiple deformable templates improves the performance further. The proposed regularization terms decreases accuracy on known shapes, while improving it on unknown shapes in all cases but one.

D	MT	$\mathcal{L}_{LS}$	$\mathcal{L}_{DP}$	Liver (K) DSC	Liver (K) NSD	Hippocampus (K) DSC	Hippocampus (K) NSD	Pancreas (K) DSC	Pancreas (K) NSD	Liver (U) DSC	Liver (U) NSD	Hippocampus (U) DSC	Hippocampus (U) NSD	Pancreas (U) DSC	Pancreas (U) NSD
✓				98.22	98.01	95.62	91.17	96.15	95.62	96.52	85.43	93.18	74.54	90.61	74.43
	✓			82.65	26.18	80.13	30.98	55.39	15.66	79.24	22.42	76.82	27.21	51.64	14.16
✓	✓			**98.58**	**98.69**	**96.42**	**94.72**	96.85	97.30	96.59	85.95	93.54	76.99	93.38	81.11
✓	✓	✓		98.53	98.46	96.28	93.33	96.81	97.11	96.45	85.58	93.62	77.01	93.01	80.35
✓	✓		✓	98.52	98.42	96.32	93.10	**96.91**	**97.77**	96.69	86.21	93.79	77.13	**94.11**	**82.04**
✓	✓	✓	✓	98.50	98.33	96.09	92.85	96.76	97.03	**96.72**	**86.90**	**93.99**	**77.47**	93.31	80.92

Table 3. **Ablation Study of *ImplicitAtlas* Design on Known (K) and Unknown (U) Shapes.** D: deformation. MT: multiple templates (m=5). $\mathcal{L}_{LS}$: regularization with Laplacian Smoothness. $\mathcal{L}_{DP}$: regularization with Deformation Penalty. The performance is evaluated in terms of the DSC (%, ↑) and NSD (%, ↑) metrics. Note that MT only (without D) denotes static matching on the learned templates.

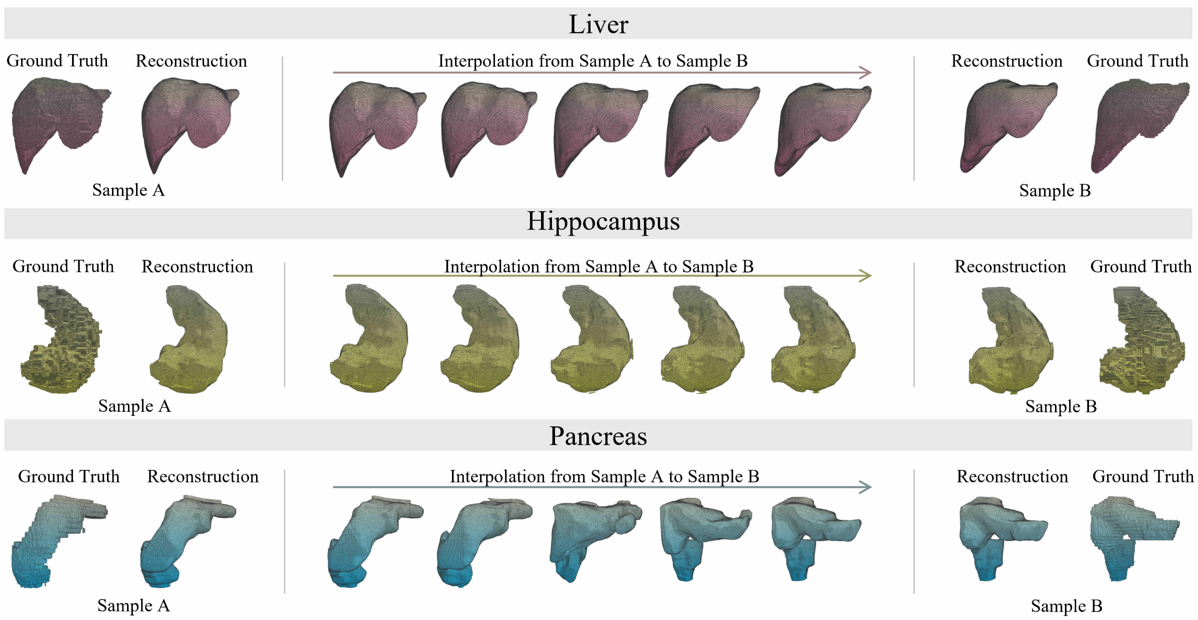

Figure 4. **Shape Reconstruction and Interpolation.** Our implicit shape representation yields reconstructions that are smoother than the annotations. Furthermore, it allows smooth interpolation from one examplar to another using linearly interpolated latent code $\mathbf{h}$.

In other words, it confers our algorithm better generalization properties.

Ablation Study. We conducted an ablation study to analyze the effectiveness of individual components of *ImplicitAtlas*. We report the results in Tab. 3. The multiple template (MT) increases the representation capacity on both known and unknown shapes, especially for more difficult cases, such as the pancreas. As observed before the regularization terms $\mathcal{L}_{LS}$ and $\mathcal{L}_{DP}$ of Eq. 7 enhance the generalization performance on unknown shapes.

Note that the single template version underperforms the pure convolutional decoder without template in some cases. For example, compare "D" in Tab. 3 to "Conv Decoder" in Tab. 2. This indicates that the performance improvement from the template is dependent on the network architectures and datasets, as previously noted in [12].

4.4. Qualitative Results

Through visualization, we now provide a qualitative analysis of *ImplicitAtlas* + reg. as defined above.

Shape Reconstruction and Interpolation. In Fig. 4, we display 2 randomly selected reconstruction per dataset. All the reconstruction are of high quality, even though the numerical metrics of Tab. 2 are not perfect and in spite of the training data artefacts depicted by Fig. 2. As our implicit shape model encodes a shape prior of the training samples, the reconstruction is much smoother than the manual annotation, which points to the potential use of our method to post-process human annotations.

We also display an interpolation between 2 samples obtained by linearly interpolating the latent codes. The interpolated samples looks valid throughout. Furthermore, the model even captures the poses, as can be seen in the bottom

Method	Liver (K5)		Hippocampus (K5)		Pancreas (K5)		Liver (U)		Hippocampus (U)		Pancreas (U)	
	DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD
MLP Decoder [38, 44]	96.67	96.51	93.52	91.31	95.20	95.83	88.33	43.90	83.62	45.07	68.97	27.70
+ Template [12, 72]	97.65	97.88	94.13	92.29	95.99	96.53	89.98	45.32	84.43	49.33	70.11	31.06
Conv Decoder [9, 45]	98.41	98.38	95.46	87.73	96.67	97.00	91.26	55.13	87.10	50.92	71.79	29.15
ImplicitAtlas	**98.89**	**99.53**	**97.02**	**96.37**	**97.23**	**98.14**	90.64	48.27	88.37	54.40	74.71	34.87
ImplicitAtlas + reg.	98.71	99.05	96.48	93.93	96.90	97.31	**92.06**	**57.69**	**89.97**	**59.39**	**81.34**	**46.78**

Table 4. **Reconstruction Accuracy of Few-Shot Learning using 5 Training Samples.** Reconstruction accuracy on the the 5 known samples (K5) and unknown samples (U) in terms of the DSC (%, ↑) and NSD (%, ↑) metrics.

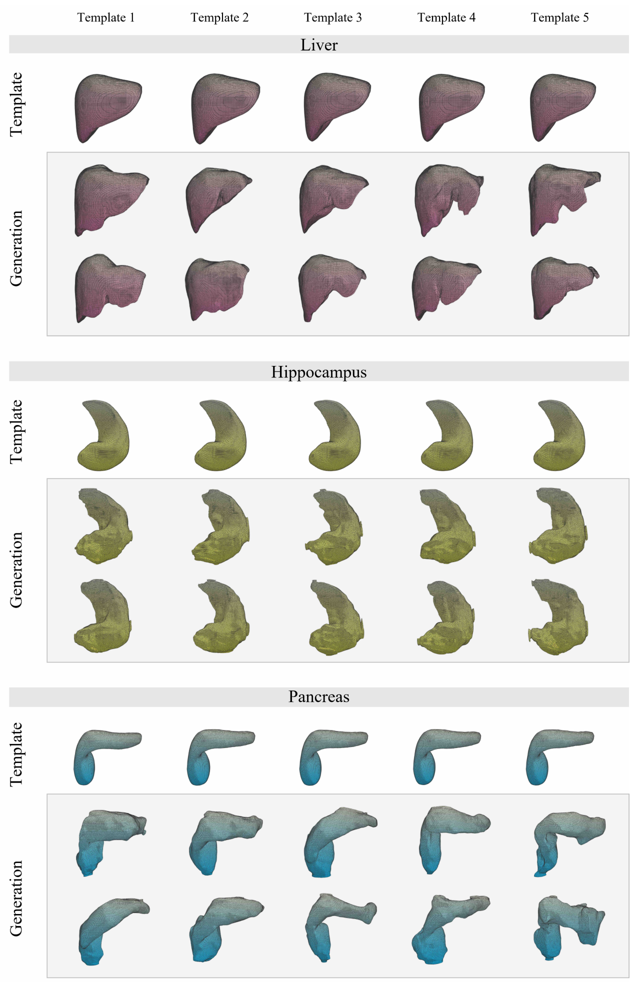

Figure 5. **Learned Templates and Shape Generation.** For each organ, we visualize the five templates we use in the top row. In the bottom rows, we show shapes obtained by sampling random latent vectors from a Gaussian prior distribution.

row of the figure. This implies that our model learns rich semantics through limited data.

Learned Templates and Shape Generation. As each row of $\mathbf{T}$ can be interpreted as a latent template vector $\mathbf{t}$, we can use $\mathcal{T}$ to visualize them by using $\mathcal{T}(\mathbf{t}, \cdot)$ to create an occupancy field. We do this for all five templates we use at the top rows of Fig. 5. In projection, the templates look similar and their average pairwise DSC (%) are 96.10, 94.89 and 94.15 for the liver, hippocampus and pancreas, respectively. Nevertheless, as shown in Tab. 3, the existence of these multiple templates improves the representation capacity for both known and unknown shapes at a negligible computational cost.

To generate new shapes, we randomly sample $\mathbf{h}$ from a Gaussian distribution, compute $\mathbf{t}(\mathbf{h})$ and $\mathbf{d}(\mathbf{h})$ of Eq. 3, and decode them into occupancy fields. Generation from a specific template can be achieved by conditioning the $\mathbf{t}$ in the forward pass. The resulting shapes are diverse and valid in most cases, as shown in the bottom rows of Fig. 5. However, there are still artifacts such as border effects and holes. More training samples and sophisticated data augmentation techniques can be expected to fix this.

4.5. Few-Shot Learning

Experiment Setting. As medical shapes are usually scarce, we consider an extreme setting on our datasets: Can we learn an effective shape model from only 5 training samples? To test this, during training, we only use the first 5 samples of the training set for each organ, while retaining all the other setting used in Sec. 4.3.

Results. We report the results in Tab. 4. Our method still significantly outperforms the baselines, especially in the more challenging cases of the pancreas. With 5 training samples, the data-hungry nature of MLP becomes obvious: It cannot provide meaningful reconstruction in some cases. The template and convolutional decoder are both useful, and the proposed regularization terms reliably improves the reconstruction accuracy.

5. Applications

The deep prior introduced by our models can be of use for numerous downstream applications in biomedical imaging. In this section, we explore some promising directions.

5.1. Shape Completion from Point Annotations

Clicking is a widespread approach to providing annotations for many biomedical applications, such as, interactive

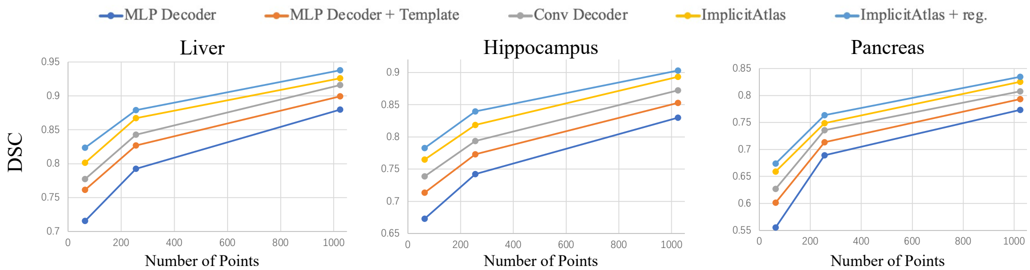

Figure 6. **Shape Completion from Point Annotations.** DSC (↑) as a function of point number (64, 256 and 1,024) on unknown shapes.

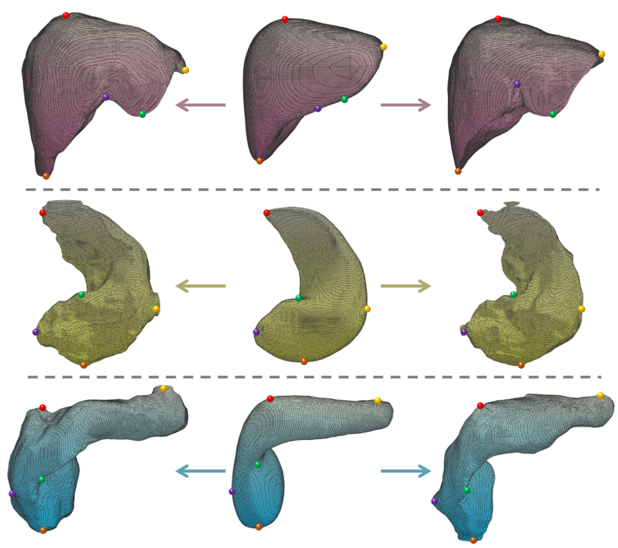

Figure 7. **Establishing Dense Correspondences.** As our method models the shapes as deformation from the learned templates (middle), correspondence can be easily established. We highlight 5 keypoints in different colors for each case.

segmentation [51], localization [57]. Here we show that *ImplicitAtlas* can generate acceptable shapes from such clicks, thus making the annotation process less cumbersome. In our experiments, we sample 64, 256 and 1,024 points close to the organ boundary of unknown shapes. A randomly initialized latent code is optimized to reconstruct the shapes by minimizing an $\mathcal{L}_{Task}$ loss to the points. We repeated these experiments using either our baselines or *ImplicitAtlas*, and plot the results in Fig. 6. The shape completion performance is evaluated by comparing the reconstructed shape with the ground truth, reported in the DSC metric. Our approach consistently outperforms the others, especially when using fewer points. We provide additional details on these experiments in the supplementary materials.

5.2. Dense Correspondences

Keypoint and landmark labeling is an important task in medical imaging [17, 47, 73]. As all the deformed points by our Implicit Deformation Network $\mathcal{D}$ are well aligned with the templates, dense correspondences between multiple shapes can be established easily. In Fig. 7, we highlight matched keypoints on the three organs. In each cases, we manually selected 5 keypoints on the template and we check where these points are sent. The resulting correspondences are visually acceptable.

The dense correspondences could be applied in many potential applications. For instance, we can create "atlas" using the implicit function: label the templates into several sub-parts, and transfer them to all the shape annotations. It also provides a way to model spatio-temporal changes, *e.g.*, tumour growth [32, 46].

6. Conclusion and Future Work

We have shown that, in spite of the challenges in biomedical applications, deep implicit surfaces could be made effective with the proposed method.

In future work, there are many extension directions. First, the current version of our method works on single organ, while many biomedical applications require multiple organs / parts. It will be particularly useful to extend *ImplicitAtlas* into a multi-class one in medical imaging area, where dealing with multiple objects in different poses and scales will be a challenge. Besides, as the implicit functions could also be used for appearance modeling, we can extend *ImplicitAtlas* to model the joint distribution of shape and appearance. A implicit model for both shape and appearance will enable a number of novel applications. Last but not least, it will be interesting to explore new applications of *ImplicitAtlas*. The deep prior encoded by the model can be directly used for improving human annotations or model predictions. It can also serve as a tool in medical image segmentation, shape analysis and multi-site generalization.

Acknowledgment. This work was supported by National Science Foundation of China (U20B2072, 61976137). This work was also supported in part by a Swiss National Science Foundation grant. We would like to thank Jan Bednařík, Benoît Guillard, and Oner Doruk for their generous help in proofreading, and the anonymous (meta-)reviewers for their valuable comments.

References

[1] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, Bram van Ginneken, et al. The medical segmentation decathlon. *arXiv Preprint*, 2021. 4

[2] A. Atzeni, M. Jansen, S. Ourselin, and J.E. Iglesias. A Probabilistic Model Combining Deep Learning and Multi-Atlas Segmentation for Semi-Automated Labelling of Histology. In *Conference on Medical Image Computing and Computer Assisted Intervention*, 2018. 2

[3] Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv Preprint*, 2013. 1, 3

[4] Jan L Bruse, Maria A Zuluaga, Abbas Khushnood, Kristin McLeod, Hopewell N Ntsinjana, Tain-Yen Hsia, Maxime Sermesant, Xavier Pennec, Andrew M Taylor, and Silvia Schievano. Detecting clinically meaningful shape clusters in medical image data: metrics analysis for hierarchical clustering applied to healthy and pathological aortic arches. *IEEE Transactions on Biomedical Engineering*, 64(10):2373–2383, 2017. 2

[5] Juan J Cerrolaza, Mirella López Picazo, Ludovic Humbert, Yoshinobu Sato, Daniel Rueckert, Miguel Ángel González Ballester, and Marius George Linguraru. Computational anatomy for multi-organ analysis in medical imaging: A review. *Medical image analysis*, 56:44–67, 2019. 2

[6] A. Chang, T. Funkhouser, L. G., P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu. Shapenet: An Information-Rich 3D Model Repository. In *arXiv Preprint*, 2015. 1

[7] Ye Chen, Jinxian Liu, Bingbing Ni, Hang Wang, Jiancheng Yang, Ning Liu, Teng Li, and Qi Tian. Shape self-correction for unsupervised point cloud understanding. In *International Conference on Computer Vision*, pages 8382–8391, 2021. 2

[8] Z. Chen and H. Zhang. Learning Implicit Fields for Generative Shape Modeling. In *Conference on Computer Vision and Pattern Recognition*, 2019. 1, 2, 3, 5

[9] J. Chibane, T. Alldieck, and G. Pons-Moll. Implicit Functions in Feature Space for 3D Shape Reconstruction and Completion. In *Conference on Computer Vision and Pattern Recognition*, 2020. 1, 3, 5, 7

[10] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 424–432, 2016. 2

[11] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 424–432. Springer, 2016. 2

[12] Yu Deng, Jiaolong Yang, and Xin Tong. Deformed implicit field: Modeling 3d shapes with learned dense correspondence. In *Conference on Computer Vision and Pattern Recognition*, pages 10286–10296, 2021. 1, 2, 3, 5, 6, 7

[13] S. Dong, G. Luo, K. Wang, S. Cao, A. Mercado, O. Shmuilovich, H. Zhang, and S. Li. VoxelAtlasGan: 3D Left Ventricle Segmentation on Echocardiography with Atlas Guided Generation and Voxel-To-Voxel Discrimination. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 622–629, 2018. 2

[14] Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. Meshcnn: a network with an edge. *ACM Transactions on Graphics*, 38(4):1–12, 2019. 2

[15] Rana Hanocka, Gal Metzer, Raja Giryes, and Daniel Cohen-Or. Point2mesh: a self-prior for deformable meshes. *ACM Transactions on Graphics*, 39(4):126–1, 2020. 2

[16] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: a review. *Medical image analysis*, 13(4):543–563, 2009. 1, 2

[17] Mattias P Heinrich, Heinz Handels, and Ivor Ja Simpson. Estimating large lung motion in copd patients by symmetric regularised correspondence fields. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 338–345. Springer, 2015. 8

[18] Ngoc-Vuong Ho, Tan Nguyen, Gia-Han Diep, Ngan Le, and Binh-Son Hua. Point-unet: A context-aware point-based neural network for volumetric segmentation. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 644–655. Springer, 2021. 2

[19] Xiaoyang Huang, Jiancheng Yang, Yanjun Wang, Ziyu Chen, Linguo Li, Teng Li, Bingbing Ni, and Wenjun Zhang. Representation-agnostic shape fields. In *International Conference on Learning Representations*, 2021. 2

[20] Y. Huo, Z. Xu, K. Aboud, P. Parvathaneni, S. Bao, C. Bermudez, S. Resnick, L.E. Cutting, and B.A. Landman. Spatially Localized Atlas Network Tiles Enables 3D Whole Brain Segmentation from Limited Data. In *Conference on Medical Image Computing and Computer Assisted Intervention*, 2018. 2

[21] Juan Eugenio Iglesias and Mert R Sabuncu. Multi-atlas segmentation of biomedical images: a survey. *Medical image analysis*, 24(1):205–219, 2015. 1, 2, 3

[22] V. Iglovikov and A. Shvets. Ternausnet: U-Net with VGG11 Encoder Pre-Trained on Imagenet for Image Segmentation. In *arXiv Preprint*, 2018. 2

[23] Eldar Insafutdinov and Alexey Dosovitskiy. Unsupervised learning of shape and pose with differentiable point clouds. In *Advances in Neural Information Processing Systems*, 2018. 2

[24] F. IsenseeEmail, J. Petersen, A. K. Zimmerer, P. Jaeger, S. Kohl, and J. Wasserthal. Nnu-Net: Self-Adapting Framework for U-Net-Based Medical Image Segmentation. In *arXiv Preprint*, 2018. 2

[25] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu. Spatial Transformer Networks. In *Advances in Neural Information Processing Systems*, pages 2017–2025, 2015. 3

[26] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. *arXiv Preprint*, 2016. 3

[27] Davood Karimi, Haoran Dou, Simon K Warfield, and Ali Gholipour. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical Image Analysis*, 65:101759, 2020. 1

[28] Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. Neural 3d mesh renderer. In *Conference on Computer Vision and Pattern Recognition*, pages 3907–3916, 2018. 2

[29] A. Kavur and M. Selver. CHAOS Challenge - Combined (CT-MR) Healthy Abdominal Organ Segmentation. In *arXiv Preprint*, 2020. 2

[30] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 3

[31] S. Koch, A. Matveev, Z. Jiang, F. Williams, A. Artemov, E. Burnaev, M. Alexa, D. Zorin, and D. Panozzo. ABC: A big CAD Model Dataset for Geometric Deep Learning. In *Conference on Computer Vision and Pattern Recognition*, pages 9601–9611, 2019. 1

[32] Yamin Li, Jiancheng Yang, Yi Xu, Jingwei Xu, Xiaodan Ye, Guangyu Tao, Xueqian Xie, and Guixue Liu. Learning tumor growth via follow-up volume prediction for lung nodules. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 508–517. Springer, 2020. 8

[33] Quande Liu, Qi Dou, and Pheng-Ann Heng. Shape-aware meta-learning for generalizing prostate mri segmentation to unseen domains. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 475–485. Springer, 2020. 2

[34] Jun Ma, Yao Zhang, Song Gu, Cheng Zhu, Cheng Ge, Yichi Zhang, Xingle An, Congcong Wang, Qiyuan Wang, Xin Liu, et al. Abdomenct-1k: Is abdominal organ segmentation a solved problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 2

[35] Chris J Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv Preprint*, 2016. 3

[36] T. Mcinerney and D. Terzopoulos. Deformable Models in Medical Image Analysis: A Survey. *Medical Image Analysis*, 1:91–108, 1996. 1, 2

[37] T. Mcinerney and D. Terzopoulos. Topology Adaptive Deformable Surfaces for Medical Image Volume Segmentation. *IEEE Transactions on Medical Imaging*, 18(10):840–850, 1999. 2

[38] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger. Occupancy Networks: Learning 3D Reconstruction in Function Space. In *Conference on Computer Vision and Pattern Recognition*, pages 4460–4470, 2019. 1, 2, 3, 5, 7

[39] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision*, pages 405–421. Springer, 2020. 2

[40] F. Milletari, N. Navab, and S.-A. Ahmadi. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *arXiv Preprint*, June 2016. 2

[41] Baptiste Nicolet, Alec Jacobson, and Wenzel Jakob. Large steps in inverse rendering of geometry. *ACM Transactions on Graphic*, 40(6), 2021. 2

[42] Michael Niemeyer and Andreas Geiger. Giraffe: Representing scenes as compositional generative neural feature fields. In *Conference on Computer Vision and Pattern Recognition*, pages 11453–11464, 2021. 2

[43] Stanislav Nikolov, Sam Blackwell, Alexei Zverovitch, Ruheena Mendes, Michelle Livne, Jeffrey De Fauw, Yojan Patel, Clemens Meyer, Harry Askham, Bernardino Romera-Paredes, et al. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. *arXiv Preprint*, 2018. 5

[44] J. J. Park, P. Florence, J. Straub, R. A. Newcombe, and S. Lovegrove. DeepSdf: Learning Continuous Signed Distance Functions for Shape Representation. In *Conference on Computer Vision and Pattern Recognition*, 2019. 1, 2, 3, 4, 5, 7

[45] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *European Conference on Computer Vision*, pages 523–540. Springer, 2020. 1, 2, 3, 5, 7

[46] Jens Petersen, Paul F Jäger, Fabian Isensee, Simon AA Kohl, Ulf Neuberger, Wolfgang Wick, Jürgen Debus, Sabine Heiland, Martin Bendszus, Philipp Kickingereder, et al. Deep probabilistic modeling of glioma growth. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 806–814. Springer, 2019. 8

[47] Maxim Pisov, Vladimir Kondratenko, Alexey Zakharov, Alexey Petraikin, Victor Gombolevskiy, Sergey Morozov, and Mikhail Belyaev. Keypoints localization for joint vertebra detection and fracture severity quantification. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 723–732. Springer, 2020. 8

[48] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Conference on Computer Vision and Pattern Recognition*, pages 652–660, 2017. 2

[49] Ashwin Raju, Shun Miao, Chi-Tung Cheng, Le Lu, Mei Han, Jing Xiao, Chien-Hung Liao, Junzhou Huang, and Adam P Harrison. Deep implicit statistical shape models for 3d medical image delineation. *arXiv Preprint*, 2021. 2

[50] Hariharan Ravishankar, Rahul Venkataramani, Sheshadri Thiruvenkadam, Prasad Sudhakar, and Vivek Vaidya. Learning and incorporating shape models for semantic segmentation. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 203–211. Springer, 2017. 2

[51] Tomas Sakinis, Fausto Milletari, Holger Roth, Panagiotis Korfiatis, Petro Kostandy, Kenneth Philbrick, Zeynettin Akkus, Ziyue Xu, Daguang Xu, and Bradley J Erickson. Interactive segmentation of medical images through fully convolutional neural networks. *arXiv preprint arXiv:1903.08205*, 2019. 8

[52] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Conference on Computer Vision and Pattern Recognition*, pages 1874–1883, 2016. 3

[53] A. Shvets, A. Rakhlin, A. Kalinin, and V. Iglovikov. Automatic Instrument Segmentation in Robot-Assisted Surgery Using Deep Learning. In *arXiv Preprint*, 2018. 2

[54] H. Spitzer, K. Amunts, S. Harmeling, and T. Dickscheid. Parcellation of Visual Cortex on High-Resolution Histological Brain Sections Using Convolutional Neural Networks. In *International Symposium on Biomedical Imaging*, pages 920–923, 2017. 2

[55] Jesse Sun, Fatemeh Darbehani, Mark Zaidi, and Bo Wang. Saunet: Shape attentive u-net for interpretable medical image segmentation. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 797–806. Springer, 2020. 2

[56] Nima Tajbakhsh, Laura Jeyaseelan, Qian Li, Jeffrey N Chiang, Zhihao Wu, and Xiaowei Ding. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, 63:101693, 2020. 1

[57] Youbao Tang, Ke Yan, Jing Xiao, and Ronald M Summers. One click lesion recist measurement and segmentation on ct scans. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 573–583. Springer, 2020. 8

[58] Hongzhi Wang, Jung W Suh, Sandhitsu R Das, John B Pluta, Caryne Craige, and Paul A Yushkevich. Multi-atlas segmentation with joint label fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):611–623, 2012. 1, 2, 3

[59] Yan Wang, Xu Wei, Fengze Liu, Jieneng Chen, Yuyin Zhou, Wei Shen, Elliot K Fishman, and Alan L Yuille. Deep distance transform for tubular structure segmentation in ct scans. In *Conference on Computer Vision and Pattern Recognition*, pages 3833–3842, 2020. 2

[60] U. Wickramasinghe, G. Knott, and P. Fua. Probabilistic atlases to enforce topological constraints. In *Conference on Medical Image Computing and Computer Assisted Intervention*, 2019. 2

[61] U. Wickramasinghe, G. Knott, and P. Fua. Deep active surface models. In *Conference on Computer Vision and Pattern Recognition*, 2021. 2

[62] U. Wickramasinghe, E. Remelli, G. Knott, and P. Fua. Voxel2mesh: 3d mesh model generation from volumetric data. In *Conference on Medical Image Computing and Computer Assisted Intervention*, 2020. 2

[63] Yuxin Wu and Kaiming He. Group normalization. In *European Conference on Computer Vision*, pages 3–19, 2018. 3

[64] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao. 3D Shapenets: A Deep Representation for Volumetric Shapes. In *Conference on Computer Vision and Pattern Recognition*, pages 1912–1920, 2015. 2

[65] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv Preprint*, 2015. 3

[66] Jiancheng Yang, Shixuan Gu, Donglai Wei, Hanspeter Pfister, and Bingbing Ni. Ribseg dataset and strong point cloud baselines for rib segmentation from ct scans. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 611–621. Springer, 2021. 2

[67] Jiancheng Yang, Yi He, Xiaoyang Huang, Jingwei Xu, Xiaodan Ye, Guangyu Tao, and Bingbing Ni. Alignshift: bridging the gap of imaging thickness in 3d anisotropic volumes. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 562–572. Springer, 2020. 2

[68] Jiancheng Yang, Yi He, Kaiming Kuang, Zudi Lin, Hanspeter Pfister, and Bingbing Ni. Asymmetric 3d context fusion for universal lesion detection. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 571–580. Springer, 2021. 2

[69] Jiancheng Yang, Xiaoyang Huang, Yi He, Jingwei Xu, Canqian Yang, Guozheng Xu, and Bingbing Ni. Reinventing 2d convolutions for 3d images. *IEEE Journal of Biomedical and Health Informatics*, 2021. 2

[70] Jiancheng Yang, Qiang Zhang, Bingbing Ni, Linguo Li, Jinxian Liu, Mengdie Zhou, and Qi Tian. Modeling point clouds with self-attention and gumbel subset sampling. In *Conference on Computer Vision and Pattern Recognition*, pages 3323–3332, 2019. 2

[71] Xi Yang, Ding Xia, Taichi Kin, and Takeo Igarashi. Intra: 3d intracranial aneurysm dataset for deep learning. In *Conference on Computer Vision and Pattern Recognition*, pages 2656–2666, 2020. 2

[72] Zerong Zheng, Tao Yu, Qionghai Dai, and Yebin Liu. Deep implicit templates for 3d shape representation. In *Conference on Computer Vision and Pattern Recognition*, pages 1429–1439, 2021. 1, 2, 3, 5, 7

[73] Heqin Zhu, Qingsong Yao, Li Xiao, and S Kevin Zhou. You only learn once: Universal anatomical landmark detection. In *Conference on Medical Image Computing and Computer Assisted Intervention*, 2021. 8